The Spanish Journal of Psychology
2008, Vol. 11, No. 1, 275-288

ISSN 1138-7416

Prediction of University Students' Academic Achievement by Linear and Logistic Models

Maria Noel Rodríguez Ayán[1] and Maria Teresa Coello García[2]
[1]Universidad de la República (Uruguay)
[2]Universidad Complutense de Madrid

University students' academic achievement measured by means of academic progress is modeled through linear and logistic regression, employing prior achievement and demographic factors as predictors. The main aim of the present paper is to compare results yielded by both statistical procedures, in order to identify the most suitable approach in terms of goodness of fit and predictive power. Grades awarded in basic scientific courses and demographic variables were entered into the models at the first step. Two hypotheses are proposed: (a) Grades in basic courses as well as demographic factors are directly related to academic progress, and (b) Logistic regression is more appropriate than linear regression due to its higher predictive power. Results partially confirm the first prediction, as grades are positively related to progress. However, not all demographic factors considered proved to be good predictors. With regard to the second hypothesis, logistic regression was shown to be a better approach than linear regression, yielding more stable estimates with regard to the presence of ill-fitting patterns.
Keywords: logistic versus linear regression, prediction, credits, academic achievement, advance in career

Se estudia el efecto de dos tipos de factores sobre el rendimiento de estudiantes universitarios: variables académicas de rendimiento previo y variables demográficas, mediante modelos lineales y logísticos. El principal objetivo del trabajo es comparar los resultados obtenidos con ambas técnicas estadísticas, para determinar cuál de ellos es más adecuado en términos de ajuste y capacidad predictiva cuando se pretende explicar y predecir el rendimiento académico, en función de variables de rendimiento previo y factores sociodemográficos. Como medida del rendimiento a predecir se empleó el avance en la carrera. Las hipótesis planteadas son: 1) El avance está directamente relacionado con las calificaciones en materias básicas de primer año y con variables demográficas y 2) Los modelos logísticos son más adecuados que los modelos lineales, ya que presentan mayor capacidad predictiva. Los resultados permiten confirmar la primera hipótesis en su primera parte, ya que el rendimiento previo está directa y significativamente asociado al avance en la carrera. Pero se cumple de forma parcial por lo que se refiere al efecto factores demográficos. Con respecto a la segunda hipótesis, la regresión logística mostró ser más adecuada que la lineal, pues arroja estimaciones más estables en relación con la presencia de patrones de mal ajuste.
Palabras clave: regresión logística versus regresión lineal, predicción, créditos, rendimiento académico, avance en la carrera

Correspondence concerning this article should be addressed to Maria Noel Rodríguez Ayán, Facultad de Química, CC 1157, Gral. Flores 2124 CP 11800, Montevideo (Uruguay). Phone/Fax: 5982 – 929-0770. E-mails: mayan@fq.edu.uy and mnrayan@yahoo.com

Translation: Virginia Navascués Howard.

University students' achievement difficulties have been a recurring concern for Higher Education Institutions for various reasons: the generalized concept that an improvement in achievement implies a higher graduation rate (Alexander, 2000; Tinto, 1993), the increasingly frequent association of academic achievement with budget issues (Burke, Modarresi, & Serban, 1999; Nonis & Wright, 2003), the need for the universities to improve their students' achievement standards because of pressure by credit agencies, requirements of prospective employers, and competence with other universities (Nonis & Wright, 2003).

Various theoretical frameworks have been proposed to address academic achievement. Although within the same theoretical framework, the variables employed can be operationalized in different ways, which makes the measurements used in the diverse studies vary from one study to another. Consequently, the results of the investigations may be different, both with regard to the effects considered as predictors and with regard to their meaning (Goenner & Snaith, 2004). Some of the studies are limited to specific settings, whose conditions are not generalizable (Nonis & Wright, 2003).

Of the diverse factors used to determine academic achievement, the most relevant are: prior achievement, learning strategies, expectations of success, sex, psychosocial factors (Cassidy & Eachus, 2000; Makinen & Olkinuora, 2004; Van den Berg & Hofman, 2005), and more recently, factors related to the educational institutions (Van den Berg & Hofman, 2005; Yorke, 2004). Educational research has shown that prior achievement is the best predictor of future achievement (Goberna, López, & Pastor, 1987; House, Hurst, & Keely, 1996; Mathiasen, 1984; McKenzie & Schweitzer, 2001; Wilson & Hardgrave, 1995; Zeegers, 2004).

The study of the effect of the variable sex on achievement has produced contradictory results. Some works suggest the existence of differential achievement due to differences in men's and women's learning styles (Lundeberg & Diemert, 1995; Martínez, 1997). In contrast, Clifton, Perry, Adams, and Roberts (2004) found that grades were not associated with sex. Van den Berg and Hofman (2005) found that in the Masters' stage, women slightly surpassed men in university progress, although this did not occur in technological careers.

Operationalization of Academic Achievement

Grades are the most universally accepted indicators of achievement in educational settings that focus on the student (Anaya, 1999; Biggs, 1989; Goberna et al., 1987; Harackiewicz, Barron, & Elliot, 1998; Pardo & Olea, 1993). Sirin (2005) performed a meta-analysis on the relation between academic achievement in primary and secondary school and socio-economic status, of articles in the databases *Education Resources Information Center* (*ERIC*), *PsycINFO,* and *Sociological Abstracts* during 1990-2000. The indicators of achievement used in the 58 articles that met the search criteria were the students' grades in the specific domains (mathematics, sciences, and verbal aptitude), as well as the general average grade.

A less frequently employed indicator to date derives from the academic credits obtained by students during a certain period of studies, although credits would be more comparable between classes or centers than grades (Nurmi, Aunola, Salmela-Aro, & Lindroos, 2003). The credits accumulated by the students would allow the establishment of advance or progress in the career (taking into account the total credits required to graduate), but not the degree of suitability of the knowledge acquired. In an investigation of the factors that affect university students' progress, Bivin and Rooney (1999) used the tobit technique to predict the number of credits accumulated by the 1989 cohort at the end of each academic year until 1994, as well as the number of credits accumulated by all the students registered in the university during 1994. Their results showed that, in the first case, the University entrance test grade had a positive effect on the number of credits accumulated only during the year of entrance, but not in the following years. In the second case, the students' average grades in the University during 1993 had a positive significant effect on the number of credits obtained during 1994. That is, previous achievement—the entrance test grade or average grade in the previous year—had a positive effect on advance.

Nonis and Wright (2003) studied the influence of motivation, the level of optimism, and the students' aptitude on the two achievement indicators: general average grade and advance. The results show that the students with a high level of optimism and motivation obtained higher achievement, either measured by grades or by the list of credits, but the magnitude of these effects was higher for the grades. Moreover, the effect of aptitude on grades would be moderated by motivation and optimism, so that the higher the levels of optimism and motivation, the higher the effect of aptitude on achievement. The same cannot be said for the indicator based on credits, for which no significant interaction effects between aptitude and the other variables was found. That is, the two indicators are affected by the same variables, but with different intensity.

Technical Models and Statistics

There are many reviews of articles published in the educational field with the goal of classifying investigations according to their design and methodological characteristics. For instance, Goodwin and Goodwin (1985) reviewed the publications that appeared in the *American Educational Research Journal* (*AERJ*) during 1979-1983, and Elmore and Woehlke (1998) analyzed the articles of the *AERJ*, the *Educational Researcher* and the *Review of Educational Research* during 1978-1987. The general conclusion was that the quantitative techniques employed only required an intermediate level of statistical knowledge and basic psychometry to understand the articles. In a more recent review of the publications of the *AERJ* and the *Journal of Counseling Psychology* during 1988-1997, Kieffer, Reese, and Thompson (2001) found a predominance of univariate techniques compared with multivariate techniques (a ratio of 3:1 in *AERJ*), observing that the use of each one of the techniques was stable during those 10 years.

In the specific field of Higher Education, the works of Kuh, Bean, Bradley and Coomes (1986) and of Kuh, Bean, Bradley, Coomes, and Hunter (1986) on publications between 1969 and 1983 also reveal scarce use of multivariate techniques compared with the abundance of descriptive analyses, although their investigation was restricted to College students, not universities. Volkwein, Carbone, and Volkwein (1988) reviewed the articles of a prestigious research journal in Higher Education, *Research in Higher Education* (*ResHE*), during 1973-1987. They also conclude that the techniques employed were stable during that interval and that 50% of the publications report the use of some multivariate analysis technique. More recently, Hutchinson and Lovell (2004) reviewed the publications of *ResHE* and two other journals, also leaders in Higher Education research: *The Review of Higher Education* (*RevHE*) and *The Journal of Higher Education* (*JHE*), during the interval 1996-2000. They found an increase in the use of multivariate statistical techniques with regard to the reviews of Kuh et al. (1986; 1986), identifying a total of 143 multiple regressions in 252 articles that presented some type of quantitative analysis. Within the multiple regressions, 113 were linear models and 30 logistic models.

Peng, So, Stage, and St. John (2002) reviewed the articles published in the *ResHE*, *JHE*, and *RevHE* in the interval 1988-1999, finding 52 articles that used logistic regression. Among them, the majority (29) referred to issues of registering and university staying, whereas only 5 used this technique for academic achievement studies. According to Peng et al. (2002), this may be because the variables of the type registering and dropping out are typically categorical or dichotomic.

Summing up, within nonexperimental quantitative educational research, multiple linear regression has been the most frequently employed multivariate technique to predict academic achievement (De la Orden, Oliveros, Mafokozi, & González, 2001; García, Alvarado, & Jiménez, 2000).

Linear Models versus Logistic Models

Currently, in higher education research, there is a tendency to acknowledge the limitations of linear regression to explain the relations between categorical criterion variables of interest (success /failure, dropping out /remaining) and a series of continuous and categorical predictors (Peng et al., 2002).

Among the problems that linear regression may present are the assumptions on which the model is based: normality of the criterion variable, normal distribution and homocedasticity of the residuals. Such assumptions are difficult to fulfill for variables that are typical of the field of education or psychology (Micceri, 1989) and they are definitely not met when the criterion variable is dichotomic.

On the other hand, if the criterion variable is continuous and is dichotomized so it can be modeled by logistic regression, information is lost, so it is important to weigh the advantages and disadvantages of both techniques.

However, the predictive value of linear models for academic achievement seems to be scarce. Using achievement in middle school and/or scores in standardized university entrance tests as the predictor variables, linear models explain less than 20% percent of the variance in some cases (García et al., 2000), rising to moderate, around 40%, in other cases (McKenzie & Schweitzer, 2001; Pike & Saupe, 2002). The inclusion of variables from the educational institution (institution effect), as well as the use of hierarchical linear models, increases the variance accounted for but not beyond 6-7% (Pike & Saupe). The inclusion of noncognitive factors (motivation, study habits) does not improve the percentage of explained variance (Mouw & Khanna, 1993), presumably due to the strong relation between psychosocial variables and prior achievement (Noble, Davenport, Schiel, & Pommerich, 1999). In fact, in a study in which previous achievement was not included in the models, Clifton et al. (2004) practically only explained 13.8% of the variance of mean grades from the demographic variables and pedagogical factors. Although this percentage increases when psychosocial variables are added, it barely reaches the value of 23.2%.

The results described show that there is an important percentage of the variance of achievement (over 50%) that is not explained by linear models. This has in part led to the search for alternative analysis techniques that allow predicting academic success /failure, such as discriminant analysis (García, 1986; Jiménez, 1987; Kelly, Holloway, & Chapman, 1981; Remus & Wong, 1982; Wilson & Hardgrave, 1995) and, to a lesser extent, logistic regression (García et al., 2000; Wilson & Hardgrave, 1995). The latter seems to be superior to linear regression (García et al., 2000) in predictive capacity.

The statistical model of logistic regression is more flexible than linear regression because it does not require the variables to meet the aforementioned assumptions of normality and homocedasticity, it allows the criterion variable to be dichotomic or polytomic and it is useful when one does not expect the relation between the predictor variables and the criterion to be linear (Tabacnik & Fidell, 1989).

The main goal of the present work is to compare the results obtained with both statistical techniques, linear and logistic regression, to determine which of them is more adequate in terms of fit and predictive capacity when we wish to explain and predict academic achievement, defined as advance in the career, as a function of prior achievement and sociodemographic factors.

We contrasted two global hypotheses:

1. Advance in the career (measured as the proportion of accumulated credits) is directly related to: a) university students' achievement, measured by the grades obtained in three basic materials (mathematics, chemistry, and biology) and (b) demographic variables (sex, type of high school, type of institution, and location where the student studied) and, therefore, both sets of variables are good predictors of such advance.
2. Logistic regression models are more adequate than linear regression models to predict advance in the career (curricular adjustment / lagging behind—normative or theoretical) because they present a better fit to the data and have better predictive capacity.

Method

Participants

The participants were university students registered in university careers of chemistry of the University of the Republic of Uruguay during the interval 2000-2003, who were registered in December of 2004. Students who had dropped out at that time were not included.

From the data of the University files, only the students who had passed the three basic areas (biology, chemistry, and mathematics) in the first year were selected, in order to include prior achievement in the models as an explanatory variable, because we did not have the students' high school grades. In total, there were 639 participants: 74.2% were female and 25.8% male, 62.3% were from Montevideo and the rest from the inland of the country. Of the sample, 64.5% were from public middle school institutions and 31% were from medical high school. Age was distributed as follows: 41.8% between 20 and 24 years (typical university student age), 32.6% between 25 and 29 years (early extra-age) and 25.7% were over 30 years (late extra-age), with a mean age of 27.1 years.

Measurements

Criterion Variables: To assess students' achievement, we used their advance in the career, as defined below:

Advance: The ratio between the number of credits accumulated by the student to date and the number of credits that he or she should have accumulated, according to the year of entrance in the University. It varies from 0 to 1, has a mean value of 0.537 and a standard deviation of 0.21.

The variable advance was dichotomized for analysis by logistic regression using two rules, giving rise to two possible curricular situations: adjustment (adequate advance in the career for the selected dichotomization criterion) and falling behind (curricular lag).

Dichotomized Advance: The rules for dichotomization were: a) statistical (using the median as cutting point) and b) theoretical (using the value of 0.75 as cutting point). The first rule defines situations of curricular adjustment and lagging behind on the basis of the participants' general result (normative dichotomization), according to whether their advance in the career was above or below the group median. Students who were below the median were considered to have lagged behind and those who were above the median were in a situation of adjustment. The second rule defines curricular adjustment as a degree of advance between 75 and 100% of the foreseen theoretical advance (dichotomization according to the theoretical advance foreseen in the study plan). Students who did not reach 75% of advance were considered to be in a situation of curricular lag.

Explanatory Variables: As explanatory variables of achievement (advance in career), three academic variables and four demographic ones were used.

Academic Variables: The academic variables included are prior achievement: the grades obtained in the areas of biology (ABI), chemistry (ACH), and mathematics (AMA). The values of each of these grades are within the interval of 3-12 points.

Demographic Variables: We included the variables sex, location in the country where middle school was studied (Montevideo / rest of the country), type of middle school institution (public / private), and orientation chosen in high school (medicine / engineering). Students who had studied abroad or who had studied a different middle school alternative (for example, agriculture high school) were not included, as they represented less than 5% of the student population.

Design

This investigation used a descriptive, correlational, and cross-sectional design, with no experimental manipulation of the variables. In order to validate the models, the total group was subdivided into two groups, Group 1 and Group 2, by stratified random sampling, to respect the distribution of the relevant nonmetric variables, and the results of both samples were compared. Group 1 was used to construct the models and Group 2 to validate them.

Data Analysis

Multivariate linear and logistic analyses were carried out. The criterion to include a variable in the predictive models, both linear and logistic, was that the statistical significance of the bivariate association between this variable and the variable to be explained was ≤ 0.25, to avoid discarding a priori variables that, in the presence of other factors in the multivariate models, could make a significant contribution ($p \leq .05$).

We used the coding method "parameterization of reference cell" to code the dichotomic variables, assigning the value (0) to the categories Female (Sex), Medicine /High School), Public (Middle School Teaching Institution) and Montevideo (Location), and the value (1) for the categories Male, Engineering, Private and Inland.

Linear Regression. Firstly, the assumptions of the model were verified. To contrast the normality, we used the Kolmogorov-Smirnov (KS) test with Lilliefors' correction. The variables that did not meet the assumption were transformed by Blom's procedure. The suitability of the models was assessed according to the significance of R^2 and the percentage of variance explained. Moreover, studies of the first- and second-order interaction and diagnostic statistics (standard residuals, Cook's distance, values of influence, standardized fit, and covariation ratio) were used to analyze the stability of the estimations compared to cases of influence (McCullagh & Nelder, 1989).

Logistic Regression. We used the two dichotomization rules indicated in the measures section. Before conducting the regression, the model assumptions were verified. Regarding linearity with the logit, the original quantitative variables ACH and ABI met the assumption, whereas AMA did not. Analysis of the indicator variables of AMA suggested a dichotomy, with the cutting point at the value of 6. Despite the fact that the practice of dichotomizing a quantitative explanatory variable can have negative effects on the results, such as loss of effect size and of statistical significance (MacCallum, Zhang, Preacher, & Rucker, 2002), in this case, the procedure is justified because the variable in its original measurement scale did not fulfill one of the prerequisites of logistic regression. Moreover, the cutting point selected, although it proceeds from statistical results, has a substantial basis. According the grading system of the University of the Republic of Uruguay, a passing grade below 6 is considered normal, whereas 6 and over correspond to good, very good, and extremely good. Therefore, dichotomization of AMA involved grouping the students according to "good" and "regular" achievement in this discipline.

In this case, when transforming quantitative variables by principal components analysis (PCA) with varimax rotation to eliminate colinearity between explanatory factors, none of the orthogonal components had a linear relation with the logit. We solved this difficulty by weighing colinearity versus not fulfilling the assumption of linearity with the logit, in order to decide which would be the best strategy.

The significance of the logistic regression parameters was performed with Wald's statistic and the change in the likelihood statistic, as well as by the chi-square statistics of Pearson's likelihood ratio (P) and Deviation (D) test and Hosmer and Lemeshow's statistic (Hosmer & Lemeshow, 1989), considering an alpha value of .05.

We also included studies of first- and second-order interaction and we estimated diagnostic statistics from configurations of cases (Hosmer & Lemeshow, 1989).

To examine the intensity of the association between each predictor variable and the criterion, we used the Odds Ratio (OR) (Agresti, 1990). For the dichotomization rules of the criterion, we considered the logit of curricular lag compared to adjustment.

Comparison of the Models. To compare the predictive efficacy of the models, the following was taken into account: the significance and magnitude of the main effects, the global predictive capacity, the percentage or variance explained, and the contribution to fit and/or to the estimations of cases with good fit and/or poor estimations.

All the analyses were performed with the SPSS 11.0 (2001) statistical package.

Results

In Table 1 are shown the descriptive statistics of the variables of the study for Groups 1 and 2.

Linear Regression

To construct the linear regression model, we used Group 1. Firstly, before the linear regression analysis, we verified the assumption of normality of the criterion variable, using the Kolmogorov-Smirnov (KS) test. The result was KS = 0.067, $p < .002$ (with Lilliefors' correction of significance), so that the hypothesis of normality was rejected. After transforming this variable by Blom's procedure, we obtained KS = 0.019, p = .200, so the hypothesis of normality of the transformed variable was accepted for an alpha value of .05.

Next, we performed the diagnosis of colinearity. As there are more than two predictor variables, we took into account the values of the variance inflation factors (VIF), the condition indexes, and the proportions of variance accounted for by each dimension, instead of the zero-order correlations between the variables to be analyzed (Pedhazur, 1997). The following criteria were used to assess the magnitude of colinearity: values of condition indexes about 30 or higher, the existence of dimensions that accounted for 90% of the variance of two or more coefficients, and values of condition indexes about 10 or higher VIF values.

The results are shown in Table 2. Although the VIF were not particularly high (between 1.51 and 1.689), PCA identified two dimensions with moderate / high condition indexes: 17.17 and 25.17.

This suggests some colinearity although it did not reach the severe colinearity limit—a value of 30 or higher (Pedhazur, 1997). No dimensions that accounted for 90% of the variance of two or more coefficients were identified (Hair, Anderson, Tatham, & Black, 1998).

Table 1
Descriptive Statistics

Variables	Category		Percentage	
			Group 1	Group 2
Sex	Female		75.0	73.4
	Male		25.0	26.6
Type of High School	Medicine		32.3	29.7
	Engineering		67.7	70.3
Type of Middle School Institution	Public		62.7	66.3
	Private		37.3	33.7
Location	Montevideo		63.9	60.7
	Inland		36.1	39.3
			M (*SD*)	
		ACH	6.29 (1.6)	6.34 (1.7)
		ABI	7.30 (1.6)	7.28 (1.6)
		AMA	7.18 (1.7)	7.20 (1.7)
			Mdn	
		Advance	0.537	0.538

* *Note*. ACH = Achievement in Chemistry, ABI = Achievement in Biology
AMA = Achievement in Mathematics.

Table 2
Linear Regression: Colinearity Diagnosis

Dimension	Eigenvalue	Condition Index	Constant	MSS	HS	ACH	ABI	AMA	Location
			Proportions of variance						
1	6.619	1.000	.00	.00	.00	.00	.00	.00	.00
2	.174	6.167	.00	.17	.00	.00	.00	.00	.27
3	8.075E-02	9.054	.00	.07	.40	.07	.07	.05	.01
4	6.191E-02	10.340	.00	.44	.39	.06	.01	.00	.21
5	3.125E-02	14.553	.00	.03	.00	.29	.04	.82	.02
6	2.244E-02	17.174	.02	.01	.02	.57	.71	.07	.08
7	1.044E-02	25.178	.98	.28	.19	.01	.18	.06	.40

Variable	VIF
MSS	1.444
HS	1.051
ACH	1.689
ABI	1.619
AMA	1.345
Location	1.422

Note. MSS = Middle School System. HS = High School. ACH = Achievement in Chemistry, ABI = Achievement in Biology, AMA = Achievement in Mathematics, VIF = Variance Inflation Factor.

The colinearity detected was solved by PCA with varimax rotation among the three indicators of prior achievement, producing three orthogonal variables. The component matrix of the PCA with varimax rotation performed on the three variables of prior achievement (ABI, AMA, and ACH) is shown in Table 3.

According to the loadings of the variables on the components, they were named Deductive Aptitudes (containing mainly the variable AMA), Inductive Aptitudes 1 (containing mainly the variables ACH) and Inductive Aptitudes 2 (ABI).

In the univariate analyses, six variables that were significantly associated with advance were identified, which were included in the subsequent multivariate analyses. They were three explanatory dichotomic factors: type of middle school institution (difference of means in the criterion variable $d = 0.37$, $t(312) = 3.155$, $p < .002$), type of high school (difference of means in the criterion variable $d = 0.25$, $t(312) = 2.1$, $p < .05$) and location in the country where they studied middle school (difference of means in the criterion $d = 0.27$, $t(312) = 2.3$, $p < .02$), and the three orthogonal components of the PCA ($r = .21$, $r = .42$, and $r = .37$) based on AMA, ACH, and ABI, respectively, all with $p < .0001$). The variable sex did not present any significant association (difference of means in the criterion variable $d = .09$, $t(312) = 0.728$, $p = .5$). The multivariate linear regression models were adjusted, first with the six above-mentioned explanatory variables and then with the effects of interaction of two variables. The most suitable model of the adjusted multiple linear regression models was the main effect model with five factors: type of institution, type of high school, and the three achievements. The variable Location had no significant effect on the presence of the other five factors ($b = .022$, $p = .547$). No significant interaction effects were found. The parameters of the model were sensitive to the presence of 28 cases identified as of influence, according to the criteria established, which increased the variance accounted for by 19%, decreasing the standard error of estimation and thus improving the predictive capacity of the model.

Table 3
Principle Component Analysis, Varimax Rotation

	X_1	X_2	X_3
ACH	.150	.952	.266
ABI	.223	.280	.934
AMA	.968	.145	.204

Note. ACH = Achievement in Chemistry, ABI = Achievement in Biology, AMA = Achievement in Mathematics.

Validation of the linear model was performed with Group 2, the results of which are shown in Table 4 (see Stevens, 2001). The criterion used was to consider the nonstandardized estimations in each group, for which purpose the respective confidence intervals of 95% were compared (Pedhazur, 1997).

In Table 4, the predictor variables and the percentage of variance accounted for were of the same order, and the coefficients of the predictors were the same as for Group 1, which was used to construct the model. The effects of the inductive aptitudes (X_2 and X_3) were of the same order and higher than those of the deductive aptitude (X_1). The regression coefficients of the model are maintained if the model is constructed exclusively with the significant effects (coefficients of .101 and .167 for type of institution and type of high school, respectively, and of .253, .403, and .433 for the three components).

Regarding the power of the linear regression technique, according to Hair et al. (1998), there should be between 15 and 20 observations for each predictor to avoid identifying nearly any relation as significant. In this study, there were approximately 60 observations (considering the analyses in each subsample). Despite the fact that the minimum value of R^2 that a test with power of .80 detects for these conditions and that type I error rate of .05 was about 4% (Hair et al., 1998), R^2 was about 50%, a value of substantial relevance.

Table 4
Validation of the Linear Model

Participants	Percentage of explained variance Adjusted R^2	Standard error of estimationion
Group 1 (Construction)	.502	0.627
Group 2 (Validation)	.533	0.650

Standardized coefficients

Participants	MSS	HS	Location	x_1	x_2	x_3
Group 1 (Construction)	0.099 *	0.165 ***	0.017	0.256 ***	0.405 ***	0.434 ***
Group 2 (Validation)	0.121 *	0.202 ***	0.002	0.283 ***	0.449 ***	0.408 ***

Note. MSS = Middle School System, HS = High School, X_1 = Deductive aptitude, X_2 = Inductive aptitude 1, X_3 = Inductive aptitude 2.
*$p < .05$. ***$p < .0001$.

Logistic Regression

To construct the logistic regression models, Group 1 was used. Firstly, we performed the colinearity diagnosis, whose results are shown in Table 5.

No condition index reached the value of 30 nor did they account for more than 90% of the variance of two or more coefficients, and the tolerance values and VIF were reasonable (Hair et al., 1998; Pedhazur, 1997). Therefore, it is concluded that there is no problem of severe multicolinearity and the present level of colinearity is accepted, because it is considered preferable to using orthogonal components that violate the assumption of linearity with the logit. Although the level of colinearity is the same as in the linear regression in this case it is not considered so important, compared to violating the assumption of linearity.

The results of the logistic models for the two dichotomization rules are shown in Table 6.

From these results, we obtained the sensitivity (proportion of students correctly predicted as lagging)—76.9% (statistical rule) and 94.4% (theoretical rule)—and the specificity (proportion of students with correctly identified curricular adjustment)—70.5% and 40.6%, respectively.

Table 5
Logistic Regression: Colinearity Diagnosis

Group 1										
		Condition		Proportions of variance						
Dimension	Eigenvalue	Index	Constant	MSS	HS	ACH	ABI	AMA-d	Location	
1	6.615	1.000	.00	.00	.00	.00	.00	.00	.00	
2	.173	6.180	.00	.18	.00	.00	.00	.00	.26	
3	7.640E-02	9.305	.00	.03	.51	.07	.08	.02	.01	
4	6.208E-02	10.322	.00	.41	.27	.12	.02	.01	.19	
5	4.167E-02	12.599	.00	.14	.01	.09	.02	.78	.11	
6	2.229E-02	17.227	.01	.00	.02	.71	.70	.05	.06	
7	9.756E-03	26.038	.98	.25	.18	.00	.17	.14	.37	

Variable	VIF
MSS	1.443
HS	1.052
ACH	1.601
ABI	1.548
AMA-d	1.074
Location	1.420

Note. MSS = Middle School System. HS = High School. ACH = Achievement in Chemistry, ABI = Achievement in Biology, AMA-d = Achievement in Mathematics, dichotomized, VIF = Variance Inflation Factor.

Table 6
Assessment of Logistic Models

Statistic	Statistical Rule	Theoretical Rule
Cox & Snell	0.291	0.252
Nagelkerke	0.389	0.397
McFadden	0.249	0.288
Test of Hosmer & Lemeshow: χ^2 (8)	5.164	10.120
χ^2 Pearson (P)	280.576	254.870
χ^2Desviation (D)	289.981	196.217
χ^2 P/ n	1.07	1
χ^2 D/ n	1.11	0.75
Percentage of correct classifications	73.7	83.5

Note. Statistical Rule = Advance dichotomized by the median. Theoretical Rule = Advance dichotomized according to the theoretical rule.
* $p < .05$.

The global goodness-of-fit statistics, Pearson and deviation with chi-square distribution, were nonsignificant. Therefore, the models selected seem to represent the data adequately. The quotient of these statistics and their degrees of freedom, with values not far from 1, points in the same direction.

Regarding the diagnosis statistics, in contrast to the linear model, neither the regression coefficients nor the predictive capacity were substantially modified by eliminating some particular covariate patterns, because the changes produced were within the errors of estimation.

Validation of the models was performed using Group 2, with the two forms of dichotomization mentioned above. The results of the validation of the logistic models are shown in Tables 7 and 8. The criterion used to compare the results in both groups was the 95% confidence interval of the OR.

For the normative lag (statistical rule) (Table 7), the effects of the variables of prior achievement were of the same order in both groups. For ACH, the OR was between 1.4 and 2 for Group 1, and between 1.2 and 1.8 for Group 2. For ABI, the OR was between 1.15 and 1.7, and between 1.3 and 2, respectively, for Groups 1 and 2. For AMA (dichotomized), the OR varied between 1.7 and 7 for Group 1, and between 1.9 and 8.5 for Group 2. Another similarity between the results is that neither the location nor the type of institution had significant effects on either of the two groups. The only difference was in the type of high school, which did not reach significance in Group 1, although it did so in Group 2 (OR between 1.3 and 4.5). The regression coefficients and the predictive capacity of the model are maintained if it is constructed exclusively with the significant

Table 7
Validation of the Logistic Model: Advance Dichotomized according to Statistical Rule (Median)

GLOBAL FIT

	Predictive capacity	Pseudo R^2 Cox & Snell	Pseudo R^2 Nagelkerke	Pseudo R^2 McFadden	χ^2 Model	χ^2 Pearson (P)	P/*df*	χ^2 Deviation	(D) D/*df*	Hosmer & Lemeshow χ^2 (8)
Group 1 (Construction)	73.7	0.291	0.389	0.249	108.854	280.576	1.07	289.981	1.11	5.164
Group 2 (Validation)	76.8	0.315	0.420	0.273	122.067	345.902	1.3	284.402	1.07	7.655

PREDICTORS

		B (Coefficient)	*SE*	Wald statistic	Exp (B)	Odds Ratio (1/Exp(B))	C.I. 95% for Exp (B) Lower	C.I. 95% for Exp (B) Higher
Group 1 (Construction)	HS [c]	–.495	.300	2.732	.609	1.642	.339	1.096
	AMA-d	–1.240	.373	11.058 **	.289	3.46	.139	.601
	ACH [e]	–.539	.111	23.631 ***	.583	1.715	.469	.725
	ABI [f]	–.349	.105	10.989 **	.706	1.416	.574	.867
	MSS [g]	–.164	.340	.232	.849	1.178	.436	1.653
	Location	–.320	.339	.893	.726	1.377	.374	1.410
	Constant	9.653	1.502	41.331	15571.141			

PREDICTORS

		B (Coefficient)	*SE*	Wald statistic	Exp (B)	Odds Ratio (1/Exp(B))	C.I. 95% for Exp (B) Lower	C.I. 95% for Exp (B) Higher
Group 2 (Construction)	HS [c]	–.886	.317	7.805 *	.412	2.427	.221	.768
	AMA-d	–1.387	.386	12.923 ***	.250	4	.117	.532
	ACH [e]	–.393	.105	14.094 ***	.675	1.481	.550	.829
	ABI [f]	–.522	.114	20.973 ***	.593	1.686	.474	.742
	MSS [g]	–.054	.339	.026	.947	1.056	.487	1.841
	Location	.445	.322	1.913	1.561	0.641	.830	2.935
	Constant	9.715	1.493	42.320	16571.955			

Note. HS = High School. AMA-d = Achievement in Mathematics, dichotomized, ACH = Achievement in Chemistry, ABI = Achievement in Biology, MSS = Middle School System.
$^*p < .05$. $^{**}p < .001$. $^{***}p < .0001$.

Table 8
Validation of the Logistic Model: Advance Dichotomized according to the Theoretical Rule

GLOBAL FIT

	Predictive capacity	Pseudo R^2 Cox & Snell	Nagelkerke	McFadden	χ^2 Model	χ^2 Pearson (P)	P/*df*	χ^2 Deviation	(D) D/*df*	Hosmer & Lemeshow χ^2 (8)
Group 1 (Construction)	83.5	0.252	0.397	0.288	91.745	254.870	1	196.217	0.75	10.120
Group 2 (Validation)	86.4	0.281	0.448	0.334	106.523	300.829	1.13	162.657	0.6	7.016

PREDICTORS

		B (Coefficient)	*SE*	Wald statistic	Exp (B)	Odds Ratio (1/Exp(B))	C.I. 95% for Exp (B) Lower	Higher
Group 1 (Construction)	HS	–.622	.388	2.574	.537	1.862	.251	1.148
	AMA-d	–2.337	1.037	5.081 *	.097	10.31	.013	.737
	ACH [e]	–.426	.132	10.342 **	.653	1.531	.504	.847
	ABI [f]	–.420	.141	8.845 *	.657	1.522	.498	.867
	MSS [g]	–1.026	.422	5.912 *	.359	2.786	.157	.820
	Location	–.675	.456	2.193	.509	1.965	.208	1.244
	Constant	15.439	2.688	32.990	5072301.59			

PREDICTORS

		B (Coefficient)	*SE*	Wald statistic	Exp (B)	Odds Ratio (1/Exp(B))	C.I. 95% for Exp (B) Lower	Higher
Group 2 (Construction)	HS	–1.460	.502	8.449 **	.232	4.31	.087	.622
	AMA-d	–1.718	.775	4.913 *	.179	5.587	.039	.820
	ACH	–.759	.148	26.398 ***	.468	2.137	.351	.626
	ABI	–.178	.145	1.521	.837	1.195	.630	1.111
	MSS	–.629	.406	2.400	.533	1.876	.241	1.181
	Location	–.043	.441	.009	.958	1.044	.404	2.272
	Constant	14.886	2.497	35.551	2917477.35			

Note. HS = High School. AMA-d = Achievement in Mathematics, dichotomized, ACH = Achievement in Chemistry, ABI = Achievement in Biology, MSS = Middle School System.
*p < .05. **p < .001. ***p < .0001.

effects (coefficients –0.561, –0.317, and –1.179 for ACH, ABI, and AMA, respectively, and 73.7% of the global predictive capacity).

In the case of lag, according to the theoretical rule (Table 8), the effect of ACH is of the same order for both group—OR between 1.2 and 2 for Group 1, and between 1.6 and 2.8 for Group 2. The effect of AMA (dichotomized) was significant in both groups, but the estimations were unstable in both cases (OR between 1.4 and 77 For Group 1, and between 1.2 and 26 for Group 2). The variable location had no significant effect for any of the groups. The differences between the groups were with regard to ABI and to the type of institution, which do not reach significant effects in Group 2, and with regard to the type of high school, which had no significant effect in Group 1 but it did so in Group 2.

In this case also, the coefficients and the predictive efficacy were maintained if the model was constructed exclusively using the significant effects (coefficient -0.736 for type of institution, - .442, -0.374, and -2.195 for the three achievements and global predictive capacity of 84.8%).

Discussion

In the three initially adjusted models, the grades in basic materials of the first year and the pre-university antecedents are good predictors of advance in the career. The scores in the resulting components of the PCA and based on ACH (X_1) and ABI (X_2) present a significant and direct relation with advance in the career, and they contributed the most to explaining advance in the linear regression analysis. The component represented by AMA (X_3) presents a significant relation with advance, but of lower intensity than the other two components.

Similar results are found when the three original variables of prior achievement in biology, mathematics, and chemistry are used in the logistic regression analysis, under either one of the two dichotomization rules.

These results are consistent with the first hypothesis proposed and are in accordance with the results of Bivin and Rooney (1999) about prediction of academic progress with tobit models.

In both the linear model and in the logistic models, an absence of relation of the variable sex with advance in the career was observed. This result is not in accordance with those found by some authors when examining learning styles of men and women (Lundeberg & Diemert, 1995; Martínez, 1997) and with the studies of Yorke (2004). Our results agree with those of Clifton et al. (2004), who found no association between achievement and sex, and those of Van den Berg and Hofman (2005), who found no differences between men and women in technological studies when considering the courses accumulated in terms of credits. Perhaps the absence of relation between sex and advance in the career observed in this investigation is specific to the field of studies considered. These results should be investigated to determine whether they are maintained in other university settings.

No significant effects for the variable location where the middle school studies were performed were found in any of the three models (the linear model and the two logistic models). This indicates that, although advance in the career is different depending on the location where one studied, in the presence of the remaining factors considered, this effect is nonsignificant.

In the linear regression model, significant effects were obtained for the type of high school. The most favorable prediction was for students who studied engineering high school compared with their counterparts from medicine. In the logistic models, this effect could not be validated. The type of middle school had a significant effect in the linear model and in the logistic model for lag with the theoretical rule, and in both cases, the most favorable prediction was for students from private middle schools. However, as with the case of high school, this effect could not be confirmed in the logistic model. There are no studies in the same educational system with which to compare our results, although Naylor and Smith (2002) found opposite differences to those found here in the English educational system.

Therefore, we can conclude that the first hypothesis is partially confirmed, as prior academic achievement is directly and significantly associated with advance in the career. But it is only partially confirmed with regard to the demographic variables, because only the effect of the type of institution was significant.

Linear Models versus Logistic Models

To compare the three models (one linear and two logistic), we considered the following: the significance and magnitude of the main effects, fit to the data, global predictive capacity, the percentage or variance explained, and the contribution of cases (or configurations) with poor fit to adjustment and/or to the estimations.

With regard to the fit of the models to the data, the linear model explains 50% of the variance of advance. Although this result is better than the one obtained in works that use grades as the indicator of achievement (17%, García et al., 2000), there remains 50% of unexplained variance, which reveals the need to include other predictor variables in the equation.

The logistic models fit the data better, as indicated by the diverse goodness-of-fit statistics: Pearson's likelihood ratio (P) and Deviation (D) test and Hosmer and Lemeshow's statistic (Hosmer & Lemeshow, 1989) and the ratio between such statistics and the corresponding degrees of freedom (Goodman, 1971, cited in Agresti, 1984). Moreover, they have a predictive capacity of between 75 and 85%, in accordance with the results of García et al. (2000), despite the fact that, in the present study, psychosocial factors were not taken into account. These results are consistent with our hypothesis 2.

As the regression models employed (linear and logistic) are different in nature, we do not have strictly equivalent measures. In linear regression, the discriminant power of the model is assessed by R^2 (percentage of variance of the criterion variable explained by the model); in logistic regression (as in discriminant analysis), the approach is different: the real status of the cases is compared with the status predicted by the model, thus obtaining the percentage of correctly predicted cases (De Maris, 2002). Although, as mentioned, these statistics are not strictly equivalent, the logistic regression correctly predicts 75% of the cases, whereas the linear regression explains 50% of the variance of the criterion variable.

Regarding the diagnostic statistics (McCullagh & Nelder, 1989), the elimination of the 28 cases identified as influence significantly improved the global fit of the linear model (R^2_c increased by 19%), as well as the magnitude of the parameters, beyond their errors of estimation. This finding cannot be contrasted with the results of other authors because in the articles we consulted, no allusion was made to diagnostic statistics. For the logistic models, in contrast, neither the global predictive capacity nor the parameter estimation change (beyond the errors of estimation) if any of the rules with poor fit is eliminated. Although in this case, logistic regression seems more robust than linear regression with regard to the cases of poor fit, this finding would require a more in-depth statistical study.

Statistical Dichotomization Criterion versus Theoretical Rule

The logistic models obtained according to the two cutting points used to dichotomize advance yield consistent results, in general terms, although the validation of the logistic model for lag according to the theoretical rule is not as satisfactory as for the normative lag. For both models, prior achievement terms were good predictors of lag, although differences were observed in the effects of the demographic variables. The fit of the models to the data and the global predictive capacity were satisfactory for both models. The values of sensitivity and specificity show that the model obtained when dichotomizing by the median adequately classifies both the adjustment situations and those of lag, whereas when dichotomizing according to the theoretical rule, the model is very effective to predict lag, but not adjustment, where the percentage of correct classifications is equivalent to random chance.

For both models, the estimations and the predictive capacity were practically unaltered by the elimination of the influence patterns.

Another common aspect of the two adjusted models is the behavior of the explanatory quantitative variables regarding the assumption of linearity with the logit. In both cases, ACH and ABI fulfill the assumption, whereas AMA does not. The effect of dichotomization of this variable (at cutting point 6) is different in both models. In the prediction of normative lag, the OR (Hosmer & Lemeshow, 1989) of dichotomized AMA varies between 1.7 and 7. In contrast, in the regression of lag according to the theoretical rule, the estimation of this factor is very unstable, and its OR varies between 1.4 and 77.

Summing up, dichotomization of the quantitative criterion variable and subsequent analysis by logistic regression does not appear to introduce bias in the estimations in general, it allows us to perform good predictions of advance in the career, and yields estimations of the effects that are stable with regard to the presence of configurations of influence.

In general terms, the second hypothesis proposed in this work is confirmed, because we can conclude that logistic regression seems more adequate than linear regression, taking into account, moreover, that it does not require the assumption of normality of the variable explained, nor of homocedasticity and normality of the residuals. The comparisons carried out suggest that the adjusted logistic model for normative lag is valid, as the results of the two groups (construction and validation) are very similar.

The results of this investigation show that academic achievement, as measured by progress in the career, is also affected by students' educational antecedents (Bivin & Rooney, 1999). Students with better grades in the first year of the faculty have less risk of curricular lag in the future, either with regard to the group norm or to the theoretical rule used in this work. Moreover, students with prior engineer-type training are more likely to progress adequately in their studies than students with prior biology-type training. The models found provide, therefore, a means to predict changes in the credits accumulated throughout the career from changes in the characteristics of the entrance cohort.

From a practical point of view and given the current concern about the efficacy of university instruction (Yorke, 2004), these results may be useful for such institutions with regard to the establishment of systems that improve the likelihood of success of the students who are at a disadvantage at the start of their university careers.

Limitations of the Work

With regard to the first goal proposed, one limitation of this work is that the participants are university students of chemistry and, therefore, they are not a representative sample of the university student population. Therefore, the factors associated with achievement should be interpreted within this context.

As noted by Cumsille and Bangdiwala (2000), the conclusions of a study are strongly linked to the analysis strategies (in this case, linear and logistic regressions) and with prior data processing before the analysis. The results obtained in this investigation correspond to the measure of academic achievement selected, and they could be different for studies using other indicators of achievement.

The criterion variable selected is a quantitative variable in its original scale, which, for the logistic analyses, was dichotomized according to two rules. Within this framework, the results were consistent. However, the use of other dichotomization criteria of the dependent variable, such as the use of categorizations in a number or ordered categories, could eventually yield different results depending on the cutting points selected and the diverse logistic techniques: binary, ordinal of accumulated probability, ratio of continuity, adjacent categories, or polytomic (Ananth & Kleinbaum, 1997; Manor, Mathews, & Power, 2000).

On the other hand, in this work, psychosocial factors (learning strategies, motivation, self-concept, etc.), which would surely increase the predictive value of the models, were not considered. They should be taken into account in future research.

References

Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

Agresti, A. (1990). *Categorial data analysis*. New York: Wiley.

Alexander, F.K. (2000). The changing face of accountability: Monitoring and assessing institutions in higher education. *Journal of Higher Education, 71*, 411-430.

Ananth, C. & Kleinbaum, D (1997). Regression models for ordinal responses: A review of methods and applications. *International Journal of Epidemiology, 26*, 1323-1333.

Anaya, G. (1999). College impact on student learning: Comparing the use of self-reported gains, standardized test scores and college grades. *Research in Higher Education, 40*, 499-526.

Biggs, J. (1989). Approaches to the enhancement of tertiary teaching. *Higher Education Research and Development, 8*, 7-25.

Bivin, D., & Rooney, P. (1999). Forecasting credit hours. *Research in Higher Education, 40*, 613-632.

Burke, J., Modarresi, S., & Serban, A. (1999). Performance: Shouldn't it count for something in state budgeting? *Change*, *31,* 16-23.

Cassidy, S., & Eachus, P. (2000). Learning style, academic belief systems, self-report student proficiency and academic achievement in higher education. *Educational Psychology, 20*, 3118-322.

Clifton, R., Perry, R., Adams, C., & Roberts, L. (2004). Faculty environments, psychological dispositions and the academic achievement of college students. *Research in Higher Education, 45*, 801-829.

Cumsille, F., & Bangdiwala, S. (2000). Categorización de variables en el análisis estadístico de datos: consecuencias sobre la interpretación de resultados. *Revista Panamericana de Salud Pública / Pan American Journal of Public Health, 8*, 348-354.

De la Orden, A., Oliveros, L., Makofozi, J., & González, C. (2001). Modelos de investigación del bajo rendimiento. *Revista Complutense de Educación, 12*, 159-178.

DeMaris, A. (2002). Explained variance in logistic regression. *Sociological Methods and Research, 31*, 27-74.

Elmore, P., & Woehlke, P. (1998). Twenty years of research methods employed in American Educational Research Journal, Educational Researcher and Review of Educational Research. Presented in the Annual Meeting of the American Educational Research Association, San Diego, California.

García, J. (1986). El análisis discriminante y su utilización en la predicción del rendimiento académico. *Revista de Educación, 280*, 229-252.

García, M., Alvarado, J., & Jiménez, V. (2000). La predicción del rendimiento académico: regresión lineal versus regresión logística. *Psicothema, 12*, 248-252.

Goberna, M., López, M., & Pastor, J. (1987). La predicción del rendimiento como criterio para el ingreso en la Universidad. *Revista de Educación*, *283*, 235-248.

Goenner, C., & Snaith, S. (2004). Accounting for model uncertainty in the prediction of university graduation rates. *Research in Higher Education, 45*, 25-41.

Goodwin, L., & Goodwin, W. (1985). Statistical techniques in ERJ articles: 1979-1983: The preparation of graduate students to read the educational research literature. *Educational Researcher, 14*, 5-11.

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). Multivariate data analysis (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Harackiewicz, J. Barron, K., & Elliot, A. (1998). Rethinking achievement goals: When are they adapted for College students and why? *Educational Psychologist, 33*, 1-21.

Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley & Sons.

House, J., Hurst, R., & Keely, E. (1996). Relationship between learner attitudes, prior achievement and performance in a General Education Course: A multi-Institutional study. *International Journal of Instructional Media, 23*, 257-271.

Hutchinson, S., & Lovell, C. (2004). A review of methodological characteristics of research published in key journals in Higher Education. *Research in Higher Education, 45*, 383-403.

Jiménez, C. (1987). Rendimiento académico en la universidad a distancia. Un estudio empírico sobre su evolución y predicción (II). *Revista de Educación, 284*, 317-347.

Kelly, E., Holloway, R., & Chapman, D. (1981). Prediction of achievement for high school students in college courses. *Journal of Educational Research, 75*, 5-15.

Kieffer, K., Reese, R., & Thompson, B. (2001). Statistical techniques employed in AERJ and ICP articles from 1988 to 1997: A methodological review. *Journal of Experimental Education*, *69*, 280-309.

Kuh, G., Bean, J., Bradley, R., & Coomes, M. (1986). Contributions of student affairs journals to the literature on college students. *Journal of College Student Personnel, 27*, 292-304.

Kuh, G., Bean, J., Bradley, R., Coomes, M., & Hunter, D. (1986). Changes in research on college students published in selected journals between 1969 and 1983. *Review of Higher Education, 9*, 177-192.

Lundeberg, M., & Diemert, S. (1995). Influence of social interaction on cognition: Connected learning in science. *Journal of Higher Education, 66*, 312-335.

MacCallum, R., Zhang, S., Preacher, K., & Rucker, D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.

Makinen, J., & Olkinuora, E. (2004). University students' situational reaction tendencies: Reflections on general study orientations, learning strategies and success. *Scandinavian Journal of Educational Research, 48*, 478-491.

Manor, O., Mathews, S., & Power, C. (2000). Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology, 29*, 149-157.

Martínez, A. (1997). Understanding and investigating female friendship's educative value. *Journal of Higher Education, 68*, 119-159.

Mathiasen, R. (1984). Producing college academic achievement: A research review. *College Student Journal, 18*, 380-386.

McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McKenzie, K., & Schweitzer, R. (2001). Who succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher Education Research & Development, 20*, 21-33.

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Mouw, J., & Kahnna, R. (1993). Prediction of academic success: A review of the literature and some recommendations. *College Student Journal, 27*, 328-336.

Naylor, R., & Smith, J. (2002). Schooling effects of subsequent university performance: Evidence for the UK university population. Coventry, UK: Department of Economics, Warwick University. Retrieved September 20, 2005, from http://www2.warwick.ac.fac/soc/economics/research/papers/twerp657.pdf.

Noble, J., Davenport, M., Schiel, J., & Pommerich, M. (1999). *Relationships between noncognitive characteristics, High School coursework and grades, and test scores of ACT-tested students.* ACT Research Report Series, 99-4. Iowa City, IA: American College Testing Program.

Nonis, S., & Wright, D. (2003). Moderating effects of achievement striving and situational optimism on the relationship between ability and performance outcomes of college students. *Research in Higher Education, 44*, 327-346.

Nurmi, J.E., Aunola, K., Salmela-Aro, K., & Lindroos, M (2003). The role of success expectation and task avoidance in academic performance and satisfaction: Three studies on antecedents, consequences and correlates. *Contemporary Educational Psychology, 28*, 59-90.

Pardo, A., & Olea, J. (1993). Desarrollo cognitivo-motivacional y rendimiento académico en segunda etapa de EGB y BUP. *Estudios de Psicología, 49*, 21-32.

Pedhazur, E. (1997). *Multiple Regression in Behavioral Research.* Fort Worth, TX: Hartcourt Brace College.

Peng, C., So, T., Stage, F., & St. John, E. (2002). The use and interpretation of logistic regression in Higher Education Journals: 1988-1999. *Research in Higher Education,43*, 259-293.

Pike, G., & Saupe, J. (2002). Does High School matter? *Research in Higher Education, 43*, 187-207.

Remus, W., & Wong, C. (1982). An evaluation of five models for the admission decision. *College Student Journal, 16*, 53-59.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review. *Review of Educational Research, 75*, 417-453.

SPSS 11.0. (2001). SPSS Manual. Regression models. Chicago, IL: SPSS.

Stevens, J.P. (2001). *Applied multivariate statistics for the social sciences*. (4th ed.) Hillsdale, N.J: Erlbaum.

Tabacnik, B.G., & Fidell, L.S. (1989). *Using multivariate statistics*. (2nd ed.). New York: Harper Collins.

Tinto, V. (1993). *Leaving college: Rethinking the causes of and cures of student attrition.* (2nd ed.), Chicago: University of Chicago Press.

Van den Berg, M.N., & Hofman, W.H.A. (2005). Student success in university education: A multimeasurement study of the impact of student and faculty factors on study progress. *Higher Education, 50*, 413-446.

Wilson, R.L., & Hardgrave, B.C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational and Psychological Measurement, 35*, 186-195.

Volkwein, J.F, Carbone, D.A., & Volkwein, E.A. (1988). Research in Higher Education: Fifteen years of scholarship. *Research in Higher Education, 28*, 271–280.

Yorke, M. (2004). Institutional research and its relevance to the performance of higher education institutions. *Journal of Higher Education Policy and Management, 26*, 141-152.

Zeegers, P. (2004). Student learning in higher education: A path analysis of academic achievement in science. *Higher Education Research and Development, 23*, 35-56.

Received May, 12, 2006
Revision received February, 28, 2007
Accepted June, 6, 2008